Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at http://about.jstor.org/participate-jstor/individuals/early-journal-content.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

Functions of Matrices.

By H. B. Phillips.

1. It is the purpose of the present paper to study the functions represented by polynomials or convergent series in a matrix or a finite number of matrices. As the work is concerned mainly with the roots of the matrices, the fundamental facts about the roots are first briefly developed.*

By a matrix of the n-th order is meant a square array of n^2 elements a_{ik}, $i, k=1, 1, \ldots, n$,

$$A=\left\| \begin{matrix} a_{11} & a_{12} \ldots & a_{1n} \\ a_{21} & a_{22} \ldots & a_{2n} \\ \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} \ldots & a_{nn} \end{matrix} \right\| = \| a_{ik} \| .$$

The matrices considered in this paper will all be of the same order.

The determinant

$$|A|=\left| \begin{matrix} a_{11} & a_{12} \ldots & a_{1n} \\ a_{21} & a_{22} \ldots & a_{2n} \\ \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} \ldots & a_{nn} \end{matrix} \right| = | a_{ik} |$$

is called the determinant of A. When this determinant is zero, A is called singular.

The sum of two matrices $A=\|a_{ik}\|$ and $B=\|b_{ik}\|$ is the matrix

$$A+B=\|a_{ik}+b_{ik}\|.$$

More generally, if λ and μ are numbers, $\lambda A+\mu B=\|\lambda a_{ik}+\mu b_{ik}\|$. A matrix is zero when and only when all its elements are zero.

The product AB of $A=\|a_{ik}\|$ and $B=\|b_{ik}\|$ is the matrix

$$AB=\|\alpha_{ik}\|, \text{ where } \alpha_{ik}=\sum_{j=1}^{n} a_{ij} b_{jk} .$$

---

* The general theory of matrices is given in Bocher's "Introduction to Higher Algebra." A very complete bibliography of literature before 1907 is given in James Byrnie Shaw's "Synopsis of Linear Associative Algebra," published by the Carnegie Institution of Washington.

The products AB and BA are not in general equal. If these products are equal, A and B are called commutative.

The products of three or more matrices are associative, that is,

$$(AB)C = A(BC) = ABC.$$

The determinant of a product of matrices is equal to the product of their determinants, for example, $|ABC| = |A| \cdot |B| \cdot |C|$.

The matrix

$$I = \left\| \begin{matrix} 1 & 0 & 0 \ldots\ldots 0 \\ 0 & 1 & 0 \ldots\ldots 0 \\ \ldots & \ldots & \ldots\ldots \\ 0 & 0 & 0 \ldots\ldots 1 \end{matrix} \right\|$$

in which the elements a_{ii} are all unity and the others zero, is called the unit matrix. It is easy to see that $AI = IA = A$.

If the determinant of A is not zero, there is a matrix A^{-1} called the reciprocal of A, such that $AA^{-1} = A^{-1}A = I$. In this case, if $AB = CD$, $B = A^{-1}CD$, *i. e.*, we can divide by a non-singular matrix.

It is to be noticed that in both multiplication and division the operation can be performed on the right or on the left. B is multiplied on the left by A if the result is AB, and on the right by A if the result is BA. Similarly, B is divided on the left by A if the result is $A^{-1}B$ and on the right if the result is BA^{-1}.

2. *Identical Equations.*—The matrices $A, B, \ldots\ldots, P$ satisfy the equation

$$A + B + \ldots\ldots + P = 0$$

if their sum is a matrix $\|\sigma_{ik}\|$ with elements σ_{ik} all zero. Cayley first observed that a matrix of the n-th order satisfies an algebraic equation of the n-th degree. This may be considered a consequence of the following theorem:

THEOREM I. *Let* $A = \|a_{ik}\|$, $B = \|b_{ik}\|$, $\ldots\ldots$, $P = \|p_{ik}\|$ *be a finite number of matrices of the n-th order and let* $\sigma_{ik} = \lambda a_{ik} + \mu b_{ik} + \ldots\ldots + \rho p_{ik}$, $\lambda, \mu, \ldots\ldots, \rho$ *being numerical parameters. If* $A', B', \ldots\ldots, P'$ *are commutative with each other and satisfy the equation*

$$AA' + BB' + \ldots\ldots + PP' = 0, \tag{1}$$

they will also satisfy the n-th degree equation

$$|a_{ik}A' + b_{ik}B' + \ldots\ldots + p_{ik}P'| = 0, \tag{2}$$

obtained by replacing $\lambda, \mu, \ldots\ldots, \rho$ *in* $|\sigma_{ik}| = 0$ *by the matrices* $A', B', \ldots\ldots, P'$, *respectively.*

This theorem enables us to replace a homogeneous linear equation with matrix coefficients by one of the n-th degree with numerical coefficients. The latter equation will usually have solutions, however, that do not satisfy the former.

To prove the theorem, let E_{ik} be the matrix of the n-th order with all its elements zero except that in the i-th row and k-th column, which is unity. It is readily seen that

$$E_{ij}E_{jk}=E_{ik}, \quad E_{ij}E_{lk}=0, \quad l\neq j. \tag{3}$$

The matrix A can be written

$$A=\sum_{i,\,k=1}^{n} a_{ik}E_{ik},$$

and so (1) is equivalent to

$$\sum_{i,\,k=1}^{n} E_{ik}(a_{ik}A'+b_{ik}B'+\ldots\ldots+p_{ik}P')=0.$$

Multiplying this equation on the left by $E_{11}, E_{12}, \ldots\ldots, E_{1n}$, respectively, and using (3), we get

$$\begin{aligned}
&E_{11}(a_{11}A'+\ldots\ldots+p_{11}P')+E_{12}(a_{12}A'+\ldots\ldots+p_{12}P')+\ldots\ldots+E_{1n}(a_{1n}A'+\ldots\ldots+p_{1n}P')=0,\\
&E_{11}(a_{21}A'+\ldots\ldots+p_{21}P')+E_{12}(a_{22}A'+\ldots\ldots+p_{22}P')+\ldots\ldots+E_{1n}(a_{2n}A'+\ldots\ldots+p_{2n}P')=0,\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots,\\
&E_{11}(a_{n1}A'+\ldots\ldots+P_{n1}P')+E_{12}(a_{n2}A'+\ldots\ldots+p_{n2}P')+\ldots\ldots+E_{1n}(a_{nn}A'+\ldots\ldots+p_{nn}P')=0.
\end{aligned}$$

Since $A', B', \ldots\ldots, P'$ are commutative, we can eliminate $E_{12}, E_{13}, \ldots\ldots, E_{1n}$ by multiplying these equations (on the right) by the cofactors of the corresponding elements in the first column of the determinant

$$|a_{ik}A'+b_{ik}B'+\ldots\ldots+p_{ik}P'| \tag{4}$$

and adding. The result is

$$E_{11}|a_{ik}A'+b_{ik}B'+\ldots\ldots+p_{ik}P'|=0. \tag{5}$$

If we expand (4) and combine terms, the result is a matrix. Equation (5) shows that all the elements in the first row of that matrix are zero. For, the product of E_{11} and any matrix $\|\alpha_{ik}\|$ has as first row $\alpha_{11}, \alpha_{12}, \ldots\ldots, \alpha_{1n}$. If then $E_{11}\|\alpha_{ik}\|$ is zero, the elements $\alpha_{11}, \alpha_{12}, \ldots\ldots, \alpha_{1n}$ must all be zero. Similarly, we could eliminate $E_{11}, E_{13}, \ldots\ldots, E_{1n}$ and so get

$$E_{12}|a_{ik}A'+b_{ik}B'+\ldots\ldots+p_{ik}P'|=0,$$

showing that all the elements in the second row of the matrix represented by (4) are zero. By continuation of this argument we conclude finally that

$$|a_{ik}A'+b_{ik}B'+\ldots\ldots+p_{ik}P'|=0.$$

3. *Characteristic Equation.*—Let I be the unit matrix. Applied to the identity $AI-IA=0$, Theorem I gives

$$\begin{vmatrix} a_{11}I-A, & a_{12}I, & \dots\dots & a_{1n}I \\ a_{21}I, & a_{22}I-A, & \dots\dots & a_{2n}I \\ \dots\dots & \dots\dots & \dots\dots & \dots\dots \\ a_{n1}I, & a_{n2}I, & \dots\dots & a_{nn}I-A \end{vmatrix}=0.$$

This is an equation of the n-th degree in A called the *characteristic* equation of A. Arranged in descending powers of A (with a change of sign if n is odd), it takes the form

$$\phi(A)=A^n+\alpha_1A^{n-1}+\dots\dots+\alpha_nI=0. \qquad (6)$$

Let $a_1, a_2, \dots\dots, a_n$ be the roots of the equation

$$\phi(r)=r^n+\alpha_1r^{n-1}+\dots\dots+\alpha_n=0.$$

Then $\phi(r)$ can be factored in the form

$$\phi(r)=(r-a_1)(r-a_2)\dots\dots(r-a_n).$$

Since this is an identity in r, the coefficients of each power of r on the two sides of the equation are equal. It will then still hold when r is replaced by A, and $a_1, \dots\dots, a_n$ by $a_1I, \dots\dots, a_nI$. Hence

$$\phi(A)=(A-a_1I)(A-a_2I)\dots\dots(A-a_nI).$$

The numbers $a_1, a_2, \dots\dots, a_n$ are called roots of the matrix A. These roots satisfy the equation

$$\phi(r)=\begin{vmatrix} a_{11}-r & a_{12} & \dots\dots a_{1n} \\ a_{21} & a_{22}-r & \dots\dots a_{2n} \\ \dots\dots & \dots\dots & \dots\dots \\ a_{n1} & a_{n2} & \dots\dots a_{nn}-r \end{vmatrix}=0.$$

This expresses that the determinant $|A-rI|$ is zero. The roots of A are therefore the values of r for which the determinant $|A-rI|$ is zero.

A matrix of the n-th order may satisfy an equation of lower than the n-th degree. The equation of lowest degree satisfied by a given matrix will, however, be unique. For, if A satisfies two equations of the m-th degree, we can eliminate A^m from them and so find an equation of lower degree satisfied by A. Let

$$\psi(A)=A^m+c_1A^{m-1}+\dots\dots+c_mI=0$$

be the equation of lowest degree satisfied by A. This is sometimes called the reduced equation for A. It is clear that

$$\psi(r)=r^m+c_1r^{m-1}+\dots\dots+c_m$$

is a factor of $\phi(r)$. For, by division, we get

$$\phi(r)=Q(r)\psi(r)+R(r),$$

where $Q(r)$ is the quotient, and $R(r)$, of degree less than m, is the remainder. Since this is an identity in r, we can replace r by A and so get

$$\phi(A)=Q(A)\psi(A)+R(A).$$

But $\phi(A)$ and $\psi(A)$ are both zero. If then $R(r)$ were not identically zero, $R(A)=0$ would be an equation of lower than the m-th degree satisfied by A. Hence $R(r)=0$ and $\phi(r)=Q(r)\psi(r)$.

This shows that all the roots of the reduced equation are roots of the characteristic equation. Conversely, all the roots of the characteristic equation satisfy the reduced equation. For, since $\psi(r)$ is a polynomial, we can factor $\psi(r)-\psi(s)$ in the form

$$\psi(r)-\psi(s)=(r-s)P(r,s),$$

where $P(r,s)$ is a polynomial in r and s. Since this is an identity we can replace s by A and r by rI. Then, since $\psi(A)=0$,

$$\psi(r)I=(rI-A)P(rI,A).$$

Equating determinants of the two sides, we get

$$[\psi(r)]^n=|rI-A|\cdot|P(rI,A)|.$$

If now r is a root of A, $|rI-A|=0$, and so $\psi(r)=0$ which was to be proved.

4. *Associated Roots.*—THEOREM II. *If $A, B, \ldots, P$ are commutative matrices with roots $a_1, a_2, \ldots, a_n$; $b_1, b_2, \ldots, b_n$, etc., those roots can be arranged in sets $a_i, b_i, \ldots, p_i$, $i=1, 2, \ldots, n$ such that, if $f(a, b, \ldots, p)$ is any polynomial in $a, b, \ldots, p$, the roots of the matrix $f(A, B, \ldots, P)$ are*

$$f(a_i, b_i, \ldots, p_i), \quad i=1, 2, \ldots, n.$$

This theorem is due to Frobenius.* The following proof is somewhat more direct than the one given by him. Let $f_1(A, B, \ldots, P)$, $f_2(A, B, \ldots, P)$, etc., be polynomials and

$$f(A, B, \ldots, P)=\sum_{i=1}^{k}\lambda_i f_i(A, B, \ldots, P), \tag{7}$$

$\lambda_1, \lambda_2, \ldots, \lambda_k$ being arbitrary numbers. Since $A, B, \ldots, P$ have characteristic equations of the n-th degree, there are only a finite number of linearly independent polynomials in those matrices. We may consider $f_i(A, B, \ldots, P)$, $i=1, 2, \ldots, k$ as those polynomials, and so by a proper choice of $\lambda_1, \lambda_2, \ldots, \lambda_k$ make $f(A, B, \ldots, P)$ equal to any given polynomial in $A, B, \ldots, P$.

---

* Sitzungsberichte Berliner Akademie (1896), p. 602.

By Taylor's theorem,

$$f(A, B, \ldots, P)-f(a_i, b_j, \ldots, p_k)I$$
$$=(A-a_iI)L+(B-b_jI)M+\ldots+(P-p_kI)R, \qquad (8)$$

where $L, M, \ldots, R$ are polynomials in $A, B, \ldots, P, a_i, b_j$, etc. Form the product

$$\Pi[f(A, B, \ldots, P)-f(a_i, b_j, \ldots, p_k)I]$$

of all the differences (8) obtained by letting $i, j, \ldots, k$ vary independently from 1 to n. This product is zero. For when the product is expanded in terms of the quantities $(A-a_iI)$, $(B-b_jI)$, etc., each term of the result not containing $(B-b_jI), \ldots, (P-p_kI)$, $j, \ldots, k$ being definitely assigned, will contain all the factors $(A-a_iI)$, $i=1, 2, \ldots, n$. Hence each term contains one or more of the products

$$\begin{aligned}&(A-a_1I)(A-a_2I)\ldots(A-a_nI),\\&(B-b_1I)(B-b_2I)\ldots(B-b_nI),\\&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots,\\&(P-p_1I)(P-p_2I)\ldots(P-p_nI).\end{aligned}$$

These all vanish because the result of equating any one of them to zero is the characteristic equation of one of the matrices. Therefore

$$\Pi[f(A, B, \ldots, P)-f(a_i, b_j, \ldots, p_k)I]=0 \qquad (9)$$

for all values of the parameters $\lambda_1, \lambda_2, \ldots, \lambda_k$.

Equation (9) is a polynomial equation in $f(A, B, \ldots, P)$. If the equation of least degree satisfied for all values of $\lambda_1, \lambda_2, \ldots, \lambda_k$ by that matrix is

$$\phi(x)=0, \qquad (10)$$

$\phi(x)$ will be a divisor of

$$\Pi[x-f(a_i, b_j, \ldots, p_k)].$$

The roots of (10), *i.e.*, the roots of $f(A, B, \ldots, P)$, are all therefore obtained by assigning proper values to $i, j, \ldots, k$ in

$$f(a_i, b_j, \ldots, p_k). \qquad (11)$$

We can therefore find n sets of constants $\alpha_i, \beta_i, \ldots, \rho_i$ (namely, the sets of values $a_i, b_j, \ldots, p_k$ in (11)) such that the roots of $f(A, B, \ldots, P)$ are

$$f(\alpha_i, \beta_i, \ldots, \rho_i), \quad i=1, 2, \ldots, n. \qquad (12)$$

In particular, if $\lambda_1, \lambda_2, \ldots, \lambda_k$ are so chosen that $f(A, B, \ldots, P)=A$, we shall have $f(\alpha_i, \beta_i, \ldots, \rho_i)=\alpha_i$. The roots of A are therefore $\alpha_1, \alpha_2, \ldots, \alpha_n$. Similarly, the roots of B are $\beta_1, \beta_2, \ldots, \beta_n$, etc. By a change of notation we can then make $\alpha_i=a_i$, $\beta_i=b_i$, etc. The roots of $f(A, B, \ldots, P)$ are then $f(a_i, b_i, \ldots, p_i)$, which was to be proved.

5. *Canonical Form.*—Two sets of associated roots $a_1, b_1, \ldots, p_1$ and $a_2, b_2, \ldots, p_2$ will be considered different unless

$$a_1=a_2,\ b_1=b_2,\ \ldots,\ p_1=p_2.$$

Let $$a_i,\ b_i,\ \ldots,\ p_i, \quad i=1, 2, \ldots, m$$

be the different sets of roots of the commutative matrices $A, B, \ldots, P$. Let $\lambda, \mu, \ldots, \rho$ be numbers not satisfying any of the equations

$$\lambda a_i+\mu b_i+\ldots+\rho p_i=\lambda a_k+\mu b_k+\ldots+\rho p_k, \quad i\neq k.$$

By Theorem II, the distinct roots of $\lambda A+\mu B+\ldots+\rho P$ are

$$\lambda a_i+\mu b_i+\ldots+\rho p_i, \quad i=1, 2, \ldots, m.$$

The factors of its characteristic equation are

$$\lambda A+\mu B+\ldots+\rho P-(\lambda a_i+\mu b_i+\ldots+\rho p_i)I \\ =\lambda(A-a_iI)+\mu(B-b_iI)+\ldots+\rho(P-p_iI),$$

and the equation of least degree satisfied for all values of $\lambda, \mu, \ldots, \rho$ is

$$\prod_{i=1}^{m}[\lambda(A-a_iI)+\mu(B-b_iI)+\ldots+\rho(P-p_iI]^{r_i}=0, \tag{13}$$

the numbers $r_1, r_2, \ldots, r_m$ being the multiplicities of the roots.

Let $$\psi_1=\prod_{i=2}^{m}[\lambda(A-a_iI)+\mu(B-b_iI)+\ldots+\rho(P-p_iI)]^{r_i}$$

and, generally, let ψ_j be the product of all the factors in (13) except

$$[\lambda(A-a_jI)+\mu(B-b_jI)+\ldots+\rho(P-p_jI)]^{r_j}.$$

It is clear that

$$\psi_i\psi_k=0, \quad i\neq k, \tag{14}$$

for the product contains all the factors in the left member of (13). Let ψ_i' be the function obtained by replacing $\lambda, \mu, \ldots, \rho$ in ψ_i by $\lambda', \mu', \ldots, \rho'$. Consider the product

$$[\lambda'(A-a_1I)+\ldots+\rho'(P-p_1I)]^{r_1}\psi_1. \tag{15}$$

When this is multiplied by ψ_1' the result is zero because

$$[\lambda'(A-a_1I)+\ldots+\rho'(P-p_1I)]^{r_1}\psi_1'=0.$$

Also the product of (15) and any one of the functions $\psi_2, \psi_3, \ldots, \psi_n$ is zero by (14). Hence

$$[\lambda'(A-a_1I)+\ldots+\rho'(P-p_1I)]^{r_1}\psi_1(\psi_1'+\psi_2+\ldots+\psi_m)=0. \tag{16}$$

The roots of the matrix

$$\psi_1'+\psi_2+\ldots+\psi_m \tag{17}$$

are obtained by replacing $A, B, \ldots\ldots, P$ by $a_i, b_i, \ldots\ldots, p_i$. When they are replaced by $a_1, b_1, \ldots\ldots, p_1$ the result is not zero because ψ_1' is not zero, whereas

$$\psi_2=\psi_3=\ldots\ldots=\psi_m=0.$$

In a similar way it is seen that the other roots of (17) are not zero. We can therefore divide (16) by (17) and so get

$$[\lambda'(A-a_1I)+\ldots\ldots+\rho'(P-p_1I)]^{r_1}\psi_1=0. \tag{18}$$

Since $\lambda', \mu', \ldots\ldots, \rho'$ are arbitrary, if (18) is expanded in powers of those parameters, each coefficient will vanish. Hence

$$(A-a_1I)^{\alpha_1}(B-b_1I)^{\beta_1}\ldots\ldots(P-p_1I)^{\rho_1}\psi_1=0$$

for all positive integral values of $\alpha_1, \beta_1, \ldots\ldots, \rho_1$ such that

$$\alpha_1+\beta_1+\ldots\ldots+\rho_1=r_1.$$

In a similar way it is shown that

$$(A-a_iI)^{\alpha_i}(B-b_iI)^{\beta_i}\ldots\ldots(P-p_iI)^{\rho_i}\psi_i=0, \tag{19}$$

for all positive integral exponents such that

$$\alpha_i+\beta_i+\ldots\ldots+\rho_i=r_i. \tag{20}$$

Let

$$\psi=\psi_1+\psi_2+\ldots\ldots+\psi_m. \tag{21}$$

As in case of (17) it is shown that none of the roots of (21) are zero. We can then define a set of functions ϕ_i by the equations

$$\phi_i=\frac{\psi_i}{\psi}, \qquad i=1, 2, \ldots\ldots, m. \tag{22}$$

It is clear that

$$\phi_1+\phi_2+\ldots\ldots+\phi_m=I. \tag{23}$$

Also, by (14),

$$\phi_i\phi_k=0, \quad i\neq k. \tag{24}$$

Multiplying (23) by ϕ_i we therefore get

$$\phi_i^2=\phi_i, \qquad i=1, 2, \ldots\ldots, m. \tag{25}$$

Let $f(a, b, \ldots\ldots, p)$ be a polynomial in $a, b, \ldots\ldots, p$ and

$$f_{\alpha,\beta,\ldots\ldots,\rho}(a, b, \ldots\ldots, p)=\frac{\partial^{\alpha+\beta+\ldots\ldots+\rho}}{\partial a^\alpha\partial b^\beta\ldots\ldots\partial p^\rho}f(a, b, \ldots\ldots, p).$$

By Taylor's theorem

$$f(a, b, \ldots\ldots, p)=f(a_i, b_i, \ldots\ldots, p_i)$$
$$+\sum_\alpha\sum_\beta\ldots\ldots\sum_\rho f_{\alpha,\beta,\ldots\ldots,\rho}(a_i, b_i, \ldots\ldots, p_i)\frac{(a-a_i)^\alpha}{\underline{|\alpha}}\ldots\ldots\frac{(p-p_i)^\rho}{\underline{|\rho}}.$$

Since this is an identity in the variables $a, b, \ldots\ldots, p$, we can replace them by $A, B, \ldots\ldots, P$. Then

$$f(A, \ldots\ldots, P)=f(a_i, \ldots\ldots, p_i)I + \sum_{\alpha}\ldots\sum_{\rho} f_{\alpha,\ldots\ldots,\rho}(a_i, \ldots\ldots, p_i)\,\frac{(A-a_iI)^\alpha}{\underline{|\alpha}}\ldots\ldots\frac{(P-p_iI)^\rho}{\underline{|\rho}}. \quad (26)$$

Let $\quad A_i=\phi_i(A-a_iI), \ldots\ldots, P_i=\phi_i(P-p_iI).$

From (25) it follows that

$$A_i^\alpha B_i^\beta \ldots\ldots P_i^\rho=\phi_i(A-a_iI)^\alpha(B-b_iI)^\beta\ldots\ldots(P-p_iI)^\rho.$$

Equations (19), (20), and (22) show that this is zero if $\alpha+\beta+\ldots\ldots+\rho \overline{>} r_i$. Finally, if we multiply (26) by ϕ_i, sum for $i=1, 2, \ldots\ldots, m$, and use (23) we get

$$f(A, \ldots\ldots, P)=\sum_{i=1}^{m}\phi_i f(a_i, \ldots\ldots, p_i)+\sum_{i=1}^{m}\sum_{\alpha}\ldots\sum_{\rho} f_{\alpha,\ldots\ldots,\rho}(a_i\ldots\ldots p_i)\,\frac{A_i^\alpha\ldots\ldots P_i^\rho}{\underline{|\alpha}\ldots\ldots\underline{|\rho}}, \quad (27)$$

the summation including powers $A_i^\alpha \ldots\ldots P_i^\rho$ for which

$$\alpha+\beta+\ldots\ldots+\rho<r_i. \quad (28)$$

Equation (27) gives a form in which any polynomial in the given commutative matrices $A, B, \ldots\ldots, P$ can be expressed. We shall refer to it as the canonical form for a function of the matrices.* Its most important property is expressed in the following theorem:

THEOREM III. *If $A, B, \ldots\ldots, P$ are commutative matrices with corresponding roots $a_i, b_i, \ldots\ldots, p_i$ and $f(a, b, \ldots\ldots, p)$ is any polynomial, the matrix $f(A, B, \ldots\ldots, P)$ can be expressed as a linear function of matrices depending only on $A, B, \ldots\ldots, P$, the coefficients in the linear functions being obtained by substituting each set of roots $a_i, b_i, \ldots\ldots, p_i$ in $f(a, b, \ldots\ldots, p)$ and in its partial derivatives of order lower than the multiplicity of the root $\lambda a_i+\mu b_i+\ldots\ldots+\rho p_i$ in the equation of least degree satisfied for all values of $\lambda, \mu, \ldots\ldots, \rho$ by $\lambda A+\mu B+\ldots\ldots+\rho P$.*

A case of particular interest is that of a polynomial such that all the coefficients $f(a_i, \ldots\ldots, p_i)$, $f_{\alpha,\ldots\ldots,\rho}(a_i, \ldots\ldots, p_i)$ in (27) vanish. Then, evidently, $f(A, B, \ldots\ldots, P)=0$. We can consider the values $a_i, b_i, \ldots\ldots, p_i$ as coordinates of a point in hyperspace. In the case considered, the function $f(a, b, \ldots\ldots, p)$ has a zero of order r_i at the point $(a_i, b_i, \ldots\ldots, p_i)$. Therefore we have proved

---

* The formula for a function of a single matrix with distinct roots was given by Sylvester, *Comptes Rendus*, Vol. XCIV (1882), p. 55. The case of a single matrix with repeated roots was given by A. Buchheim, *Phil. Mag.*, (5) 22 (1886), pp. 173–174.

THEOREM IV. *If the polynomial $f(a, b, \ldots, p)$ has at each of the points $(a_i, b_i, \ldots, p_i)$, $i=1, 2, \ldots, m$, a zero of order equal to or greater than the multiplicity of the corresponding root in the equation of least degree satisfied for all values of $\lambda, \mu, \ldots, \rho$ by $\lambda A+\mu B+\ldots+\rho P$ then $f(A, B, \ldots, P)=0$.*

If the matrices ϕ_i, $A_i^\alpha B_i^\beta \ldots P_i^\rho$ in (27) are linearly independent, Theorem IV expresses the necessary and sufficient condition that $f(A, B, \ldots, P)$ vanish. This is easily shown to be true for polynomials in a single matrix.* For polynomials in two or more matrices such may not be the case. For instance, A and B could be equal. Then A_i and B_i would be equal.

6. *Commutative Matrices not Expressible as Polynomials in the same Matrix.*—The simplest illustration of commutative matrices is furnished by polynomials in a single matrix. If

$$A=\alpha_1\phi^p+\alpha_2\phi^{p-1}+\ldots+\alpha_p I, \quad B=\beta_1\phi^q+\beta_2\phi^{q-1}+\ldots+\beta_i I,$$

obviously A and B are commutative. If it were true that any two commutative matrices could be so expressed,† by a repetition of the process, any finite number could be expressed as polynomials in the same matrix. The results of the preceding sections could then be more readily obtained by using these expressions. That such is not the case will now be shown by a simple example. Let

$$A=\begin{Vmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{Vmatrix}, \quad B=\begin{Vmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{Vmatrix}$$

By direct multiplication it is shown that

$$A^2=B^2=AB=BA=0. \tag{29}$$

Hence A and B are commutative. Suppose they are expressible as polynomials in ϕ. Since the characteristic equation for ϕ is of third degree, the expressions can be reduced to the form

$$A=\alpha_1\phi^2+\alpha_2\phi+\alpha_3 I, \quad B=\beta_1\phi^2+\beta_2\phi+\beta_3 I. \tag{30}$$

If α_1 and β_1 are both zero, these polynomials are of first degree. If not, we form the expression

$$(B_1A-\alpha_1B)^2=[B_1(\alpha_2\phi+\alpha_3 I)-\alpha_1(\beta_2\phi+\beta_3 I)]^2. \tag{31}$$

Equation (29) shows that this is zero. The expression

$$[\beta_1(\alpha_2x+\alpha_3)-\alpha_1(\beta_2x+\beta_3)]^2$$

---

* See Metzler, AMERICAN JOURNAL, Vol. XIV (1892), p. 339.

† Frobenius raised this question in the article to which we have previously referred, but stated that he had not decided whether it could be done or not.

is not identically zero, however, because A and B do not satisfy an equation of the form $\beta_1 A - \alpha_1 B = 0$. By the use of (31) we can then reduce (30) to the form

$$A = a_1\phi + a_2 I, \quad B = b_1\phi + b_2 I. \tag{32}$$

The coefficients a_1 and b_1 can not be zero because A and B are not multiples of the unit matrix. Hence, as before, we form the expression

$$(b_1 A - a_1 B)^2 = (b_1 a_2 - a_1 b_2)^2 I = 0.$$

This shows that $b_1 a_2 - a_1 b_2 = 0$, and so, from (32), $b_1 A - a_1 B = 0$. Since A is not a multiple of B this is impossible. Therefore A and B cannot be expressed as polynomials in ϕ.

7. *Limits and Convergence.*—A variable matrix $\phi = \|r_{ik}\|$ is said to approach a matrix $A = \|a_{ik}\|$ as limit if

$$\text{Lim } r_{ik} = a_{ik}, \qquad i, k = 1, 2, \ldots, n.$$

In this case it is clear that each root of ϕ will approach a root of A as limit. For the roots of ϕ and A satisfy the equations

$$|rI - \phi| = r^n + \rho_1 r^{n-1} + \ldots + \rho_n = 0, \tag{33}$$

$$|aI - A| = a^n + \alpha_1 a^{n-1} + \ldots + \alpha_n = 0. \tag{34}$$

The coefficients $\rho_1, \rho_2, \ldots, \rho_n$ are definite rational integral functions of the elements r_{ik} of ϕ, and the coefficients $\alpha_1, \alpha_2, \ldots, \alpha_n$ are the same functions of the elements of A. When ϕ approaches A as limit

$$\text{Lim } \rho_i = \alpha_i, \qquad i = 1, 2, \ldots, n.$$

Also, if the coefficient of the highest power (in this case unity) does not approach zero, the roots of a polynomial equation are continuous functions of its coefficients.* Hence as ϕ approaches A as limit the roots of (33) approach those of (34) as limits.

Since different matrices can have the same roots, the roots of ϕ can approach those of A when ϕ does not approach A as limit.

The sum $\phi_1 + \phi_2 + \ldots + \phi_m$ of m matrices is equal to a single matrix S_m If S_m approaches a limit S when m increases indefinitely, the series

$$\phi_1 + \phi_2 + \ldots + \phi_m + \ldots$$

is said to converge and have the sum S.† In case of a multiple series, such as

$$\Sigma\phi_{i,j,k}, \qquad i, j, k = 1, 2, \ldots, \infty,$$

---

* Weber's Algebra, Vol. I, § 44.

† Peano treated the convergence of a matrix series by means of a modulus, *Math. Annalen* (1888), Vol. XXXII, pp. 450–456. See also G. A. Bliss, *Annals of Mathematics* (2) 6, pp. 49–68.

an order of summation is assumed to be included in the definition of the series. It is thus equivalent to a simple series.

If $A, B, \ldots, P$ are commutative matrices, by Theorem III, any polynomial $f(A, B, \ldots, P)$ is expressible as a linear function of matrices depending only on $A, B, \ldots, P$, the coefficients being $f(a_i, b_i, \ldots, p_i)$ and perhaps some of the derivatives $f\alpha, \beta, \ldots, \rho(a_i, b_i, \ldots, p_i)$. If these coefficients approach definite limits when the number of terms in the polynomial increases indefinitely, the matrix $f(A, B, \ldots, P)$ will then approach a definite limit.

THEOREM V. *If $A, B, \ldots, P$ are commutative matrices and $f(a, b, \ldots, p)$ an infinite series, the series $f(A, B, \ldots, P)$ will converge and be represented by* (27) *if the series $f(a_i, b_i, \ldots, p_i)$, $f_{\alpha, \beta, \ldots, \rho}(a_i, b_i, \ldots, p_i)$ in the right member of* (27) *converge.*

In this theorem it is understood that $f_{\alpha, \beta, \ldots, \rho}(a, b, \ldots, p)$ is obtained by differentiating $f(a, b, \ldots, p)$ term by term and arranging the results in the same order as the corresponding terms in $f(a, b, \ldots, p)$.

If $f(a, b, \ldots, p)$ is an analytic function of the variables $a, b, \ldots, p$ and $A, B, \ldots, P$ an equal number of commutative matrices, we define $f(A, B, \ldots, P)$ as the value (if it is definite) determined by (27). In order that the value be definite it is sufficient that $f(a, b, \ldots, p)$ be analytic near each of the points $(a_i, b_i, \ldots, p_i)$ determined by the distinct sets of associated roots. For then the derivatives in (27) are also definite. If the function is many valued, a definite branch must be chosen at each of the points $(a_i, b_i, \ldots, p_i)$. The same branch need not however be used for all.

8. *Taylor's Series.*—Let $f(z)$ be a function of the complex variable z analytic in the neighborhood of each root of a matrix A. If we choose a definite branch of $f(z)$ at each root (but not necessarily the same branch for different roots) the functions $f(A)$, $f'(A)$, $f''(A)$, etc. are determined by (27). Let Z be a matrix commutative with A and consider the series

$$\psi(A, Z) = f(A) + f'(A)(Z-A) + \ldots + f^m(A)\frac{(Z-A)^m}{\underline{|m}} + \ldots \quad (35)$$

Let z_i and a_i be corresponding roots of Z and A. The series (35) will converge if each root z_i lies within the circle of convergence of $f(z)$ with center at a_i. For then the series

$$f(a_i) + f'(a_i)(z_i-a_i) + \ldots + f^m(a_i)\frac{(z_i-a_i)^m}{\underline{|m}} + \ldots, \quad (36)$$

and the series obtained by substituting a_i and z_i in the partial derivatives of $\psi(x, y)$ will converge. Hence the conditions of Theorem V are satisfied.

Conversely, if (35) converges, (36) will converge, and so z_i will lie within or on the circle of convergence of $f(z)$ with center a_i. For the sum of m terms in (36) represents a root of the sum of corresponding terms in (35), and, if a matrix approaches a limit, we have shown that its roots must approach definite limits. Furthermore

$$\psi(a_i, z_i) = f(a_i) + f'(a_i)(z_i - a_i) + \ldots . + f^m(a_i)\frac{(z_i - a_i)^m}{\underline{|m}} + \ldots . = f(z_i).$$

Also, if

$$\psi_{\alpha,\beta}(x, y) = \frac{\partial^{\alpha+\beta}\psi(x, y)}{\partial x^\alpha \partial y^\beta}$$

it is readily seen that

$$\psi_{\alpha,0}(a_i, z_i) = f^\alpha(z_i), \quad \psi_{\alpha,\beta}(a_i, z_i) = 0, \quad \beta \neq 0.$$

Hence (27) gives the same expression for $\psi(A, Z)$ and for $f(Z)$. Therefore

$$f(Z) = f(A) + f'(A)(Z - A) + \ldots . + f^m(A)\frac{(Z - A)^m}{\underline{|m}} + \ldots .. \tag{37}$$

We may call this the expansion of $f(Z)$ in the neighborhood of the matrix A.

THEOREM VI. *The Taylor's series* (37) *for* $f(Z)$ *is valid for any matrix* Z *commutative with* A *if each root of* Z *lies within a circle with center at the corresponding root of* A, *in which* $f(z)$ *is analytic.*

This theorem enables us to determine an analytic function of a matrix Z by means of a power series

$$c_1 + c_2(Z - A) + \ldots . + c_m(Z - A)^m + \ldots .,$$

and its analytic continuations just as we determine an analytic function of the complex variable z by the series

$$c_1 + c_2(z - a) + \ldots . + c_m(z - a)^m + \ldots .,$$

and its analytic continuations. The principal difference in the two cases is that Z and A must be commutative and that the region of convergence of the matrix series consists of m circles (one for each distinct root of A) instead of one. On a range of commutative matrices the function $f(Z)$ is one valued if each root of Z is restricted to a region of the complex plane (not necessarily the same for different roots) in which the corresponding branch of $f(z)$ is one valued.